

Bioinformatics for Cancer Epidemiologic Research

Deborah Winn

Division of Cancer Control and
Population Sciences Health
Informatics Initiative

Primary goal of the NCI DCCPS health informatics initiative



DIVISION OF
CANCER CONTROL
& POPULATION SCIENCES

- ◎ To transform research in cancer control and population sciences by using the benefits of health informatics to
 - Improve the efficiency and reduce the difficulties of conducting research
 - Increase cross-disciplinary and collaborative research, especially through the use of standards and common platforms
 - Create opportunities for data and other linkages
 - Increase capacity for research and the use of research to understand mechanisms of cancer causation, monitor the burden of cancer, develop interventions to prevent cancer, detect it earlier, and reduce its consequences, and inform policy decisions



Trends influencing the need for a bioinformatics initiative

- ◎ NIH data sharing policy
- ◎ Collaboration occurring among scientists who may “speak different languages”
- ◎ Growth of consortia
- ◎ Need for “cross-walks” across existing studies
- ◎ Epidemiology studies involve data from many content domain types: genotypes, biochemistries, clinical record, questionnaire data

Trends influencing the need for a bioinformatics initiative (cont.)

- ◎ Constant need to do research faster, better, cheaper
- ◎ Need to translate from biology to epidemiology to prevention and back
- ◎ NIH roadmap: Bioinformatics and Computational Biology
- ◎ Strong NCI bioinformatics development efforts in other areas, e.g., HAPMAP

Scope of DCCPS Health Informatics Initiative

- ◎ Surveillance systems
- ◎ Behavioral research
- ◎ Intervention research
- ◎ Cancer outcomes and quality of care
- ◎ Cancer survivorship

NCI partners

- ◎ Division of Cancer Control and Population Sciences (DCCPS)(Extramural)
- ◎ Division of Cancer Epidemiology and Genetics (DCEG)(Intramural)
- ◎ Center for Bioinformatics (NCICB)

DCCPS and DCEG joint activities



DIVISION OF
CANCER CONTROL
& POPULATION SCIENCES

- ◎ **Development of a Vocabulary Enterprise System for population sciences and cancer control concepts**
- ◎ **A population sciences and cancer control “workspace” within the NCICB caBIG project**
- ◎ **Models and applications, e.g., risk prediction workshop**



Selected major initiatives of the NCI Center for Bioinformatics



DIVISION OF
CANCER CONTROL
& POPULATION SCIENCES

©caBIG

©Enterprise vocabulary systems



caBIG

- © An initiative of the NCICB has as its purpose a common, extensible informatics platform that integrates diverse data types and supports interoperable analytic tools. Will allow research groups to tap into the rich collection of emerging cancer research data while supporting their individual investigations. Workspaces within the caBIG initiative are areas or virtual environments where caBIG pilot project activities will be grouped.
- © DCCPS and DCEG are working towards implementation of a population sciences and cancer control workspace within caBIG to enhance both the research underway or supported by the two Divisions as well as that of other scientists engaged in cancer research

Enterprise vocabulary system

The NCI EVS is set of services and resources that address NCI's needs for controlled vocabulary. It includes activities related to common data elements and the NCI Thesaurus and Metathesaurus, which is a biomedical thesaurus created specifically to meet the needs of the NCI, is produced by the NCI EVS project.

Common Data Elements

The Common Data Elements (CDE) Dictionary contains a repository of detailed information pertaining to each standard element used to collect essential data related to cancer research

Priority content domains

- ◎ Tobacco use
- ◎ Alcohol use
- ◎ Diet
- ◎ Physical activity
- ◎ Body mass
- ◎ Medications
- ◎ Reproductive factors
- ◎ Cancer screening

Nearly completed Common Data Elements



DIVISION OF
CANCER CONTROL
& POPULATION SCIENCES

©Age

©Sex

©Race



Value of common data elements and an established caBIG workspace

- Have available a repository of information about the data in studies
- Can facilitate creation of questionnaires and forms
- Can compare data elements from different studies
- Can cross-walk between studies if they have been added to the Enterprise Vocabulary System

Value of common data elements & an established caBIG workspace (cont.)

- **Can benefit from common data elements and thesauruses from other disciplines**
- **Helps ensure that other disciplines can benefit from our knowledge about appropriate ways to collect information related to population sciences and cancer control**

Cancer Family Registries Informatics Center



DIVISION OF
CANCER CONTROL
& POPULATION SCIENCES

- ◎ **Contract to be let soon**
- ◎ **Intended to make all Cancer Family Registry informatics practices consistent with NCI Center for Bioinformatics standards**
- ◎ **Converts all of existing cancer family registry questionnaires and forms (e.g., family history information) into common data elements**
- ◎ **Will serve as a model for related efforts**



Uncle Sam
wants you!



Uncle Sam
wants your
questionnaires!

- In an advisory capacity
- On common data elements working groups
- Electronic format if possible
- Any cross-walking you have done

Some useful NCI Center for Bioinformatics websites

- © Home page <http://ncicb.nci.nih.gov/>
- © Cancer data standards repository
<http://ncicb.nci.nih.gov/core/caDSR>
- © Common data elements browser:
<http://ncicb.nci.nih.gov/CDEBrowse>
[r/](http://ncicb.nci.nih.gov/CDEBrowse) (I used smok* as my search word)
- © NCI metathesaurus page:
<http://ncimeta.nci.nih.gov/indexMeta>
[phrase.html](http://ncimeta.nci.nih.gov/indexMeta) Click on “browse” to
start.

Epidemiology and Genetics Research Program contact



DIVISION OF
CANCER CONTROL
& POPULATION SCIENCES

**Deborah M. Winn, Ph.D., Chief
Clinical & Genetic Epidemiology Research
Branch
Epidemiology and Genetics Research Program
Division of Cancer Control and Population
Sciences
Executive Plaza North, Room 5134, MSC 7393
6130 Executive Boulevard
Bethesda, MD 20892-7393 (Regular mail)
Rockville, MD 20852
301-594-9499 (Voice)
301-496-9600 (Main group number)
301-435-5477 (Fax)**





CDE Browser



CDE
Cart



Home



FormBuilder



Help

Data Element Search

Note: Enter/select search criteria and click search button to initiate search. The wildcard character is *. Click the Help button above for more information on CDEBrowser.

Search For:

Alternate Name:

Search Field(s):

- ALL
- Preferred Name
- Long Name
- Document Text

Alternate Name Type(s):

- ALL
- ABBREVIATION
- C3D Name

Permissible Value:

Value Domain:



Public ID:

Data Element Concept:



Classification:



Version: Latest Version ☒ All Versions ☐

Context Use:

Workflow Status:

- ALL
- APPRVD FOR TRIAL USE
- CMTE APPROVED
- CMTE SUBMTD

Registration Status:

- ALL
- Application
- Candidate
- Qualified

Search Data Elements

Clear

New Search

[\[Download Data Elements to Excel\]](#)

[\[Download Data Elements as XML\]](#)

Add to CDE Cart

1 - 21 of 21

| <input type="checkbox"/> | Long Name | Document Text | Owned By | Used By Context | Registration Status | Workflow Status | Public ID | Version |
|--------------------------|---|---|----------|-----------------|---------------------|----------------------|-----------|---------|
| <input type="checkbox"/> | Smoked at least 100 cigarette (5 packs) during lifetime | Smoked at least 100 cigarette (5 packs) during lifetime | CTEP | SPOREs | Qualified | APPRVD FOR TRIAL USE | 2001713 | 3.0 |
| <input type="checkbox"/> | Smoking Current Use Daily Frequency | Number of Cigarettes per day | CTEP | | Qualified | RELEASED | 2179794 | 1.0 |
| <input type="checkbox"/> | Smoking Current Use Ind-2 | Does the patient currently smoke | CTEP | | Qualified | RELEASED | 2180053 | 1.0 |



CDE Browser



Back



Help

Data Element

Data Element Concept

Permissible Values

Classifications

Usage

Data Element Derivation

Data Element Details

| | |
|------------------------------|---|
| Public ID: | 2001713 |
| Preferred Name: | SMOKED_AT_LEAST_100_CIGARETTE_ |
| Long Name: | Smoked at least 100 cigarette (5 packs) during lifetime |
| Document Text: | Smoked at least 100 cigarette (5 packs) during lifetime |
| Definition: | Smoked at least 100 cigarette (5 packs) during lifetime |
| Value Domain: | SMOKED_AT_LEAST_100_CIGARETTE_ |
| Data Element Concept: | SMOKED_AT_LEAST_100_CIGARETTE_ |
| Context: | CTEP |
| Workflow Status: | APPRVD FOR TRIAL USE |
| Version: | 3.0 |
| Origin: | |
| Registration Status: | Qualified |

Reference Documents

| Document Name | Document Type | Document Text | URL |
|---------------|-------------------------|---|-----|
| CRF Text | LONG_NAME | Smoked at least 100 cigarette (5 packs) during lifetime | |
| CRF Text | HISTORIC SHORT CDE NAME | Smoked at least 100 cigarette (5 packs) during lifetime | |

Designations

| Name | Type | Context | Language |
|--------------------------------|---------|---------|----------|
| SMOKED_AT_LEAST_100_CIGARETTE_ | USED_BY | SPOREs | |



CDE Browser



Back



Help

[Data Element](#) [Data Element Concept](#) [Permissible Values](#) [Classifications](#) [Usage](#) [Data Element Derivation](#)

Selected Data Element

| | |
|------------------|---|
| Public ID: | 2001713 |
| Preferred Name: | SMOKED_AT_LEAST_100_CIGARETTE_ |
| Long Name: | Smoked at least 100 cigarette (5 packs) during lifetime |
| Document Text: | Smoked at least 100 cigarette (5 packs) during lifetime |
| Definition: | Smoked at least 100 cigarette (5 packs) during lifetime |
| Workflow Status: | APPRVD FOR TRIAL USE |
| Version: | 3.0 |

Value Domain Details

| | |
|-----------------------------------|---|
| Public ID: | 2016323 |
| Preferred Name: | SMOKED_AT_LEAST_100_CIGARETTE_ |
| Long Name: | Smoked at least 100 cigarette (5 packs) during lifetime |
| Definition: | Smoked at least 100 cigarette (5 packs) during lifetime |
| Workflow Status: | DRAFT NEW |
| Version: | 3.0 |
| Datatype: | ALPHANUMERIC |
| Unit of Measure: | |
| Display Format: | |
| Maximum Length: | 255 |
| Minimum Length: | |
| Decimal Place: | |
| High Value: | |
| Low Value: | |
| Value Domain Type: | Enumerated |
| Conceptual Domain Preferred Name: | CTEP |
| Conceptual Domain Context Name: | CTEP |
| Conceptual Domain Version: | 2.31 |
| Origin: | |

Permissible Values

1 - 3 of 3

| Value | Value Meaning | Value Meaning Description |
|----------------------|----------------------|---------------------------|
| No | NO | NO |
| Yes, but quit | YES, BUT QUIT | |
| Yes, currently smoke | YES, CURRENTLY SMOKE | |



CDE Browser



Back



Help

Data Element

Data Element Concept

Permissible Values

Classifications

Usage

Data Element Derivation

Selected Data Element

| | |
|-------------------------|---|
| Public ID: | 2001713 |
| Preferred Name: | SMOKED_AT_LEAST_100_CIGARETTE_ |
| Long Name: | Smoked at least 100 cigarette (5 packs) during lifetime |
| Document Text: | Smoked at least 100 cigarette (5 packs) during lifetime |
| Definition: | Smoked at least 100 cigarette (5 packs) during lifetime |
| Workflow Status: | APPRVD FOR TRIAL USE |
| Version: | 3.0 |

Classifications

[*CS:Classification Scheme](#) [CSI:Classification Scheme Item](#)

| CS* Preferred Name | CS* Definition | CSI* Name | CSI* Type |
|--|----------------|-----------|-----------|
| There are no classifications for the selected CDE. | | | |

User: Public User

Version 2.1.1



CDE Browser



Back



Help

Data Element

Data Element Concept

Permissible Values

Classifications

Usage

Data Element Derivation

Selected Data Element

| | |
|-------------------------|---|
| Public ID: | 2001713 |
| Preferred Name: | SMOKED_AT_LEAST_100_CIGARETTE_ |
| Long Name: | Smoked at least 100 cigarette (5 packs) during lifetime |
| Document Text: | Smoked at least 100 cigarette (5 packs) during lifetime |
| Definition: | Smoked at least 100 cigarette (5 packs) during lifetime |
| Workflow Status: | APPRVD FOR TRIAL USE |
| Version: | 3.0 |

Usage

1 - 4 of 4

| Protocol Number | Lead Org | Used By | Usage Type |
|-----------------|----------|---|------------|
| Iloprost | LCBCC | SMOKING HISTORY | CRF |
| RTOG P-0011 | RTOG | A5 Radiation Therapy Oncology Group Demographic Data Form - Version 2 | CRF |
| RTOG-0320 | RTOG | A5 RTOG Phase III - Multiple Brain Mets Demographic Data Form - version 2 | CRF |
| RTOG0214 | RTOG | Radiation Therapy Oncology Group Demographic Data Form - Version 2 | CRF |



CDE Browser



Back



Help

Data Element

Data Element Concept

Permissible Values

Classifications

Usage

Data Element Derivation

Selected Data Element

| | |
|-------------------------|---|
| Public ID: | 2001713 |
| Preferred Name: | SMOKED_AT_LEAST_100_CIGARETTE_ |
| Long Name: | Smoked at least 100 cigarette (5 packs) during lifetime |
| Document Text: | Smoked at least 100 cigarette (5 packs) during lifetime |
| Definition: | Smoked at least 100 cigarette (5 packs) during lifetime |
| Workflow Status: | APPRVD FOR TRIAL USE |
| Version: | 3.0 |

Data Element Derivation Details

Selected CDE is not a derived data element.

epidemiological

Search

The basic search is enabled. Click on the "Advanced" tab to customize your search criteria. You can mouse-over each advanced search item for help in utilizing the advanced features.

About
Browse
Copyright
New Term
Sources
User's Guide

Matching Concepts

| | |
|--|---|
| Epidemiological Approaches | Research Activity |
| epidemiological bias (Bias (Epidemiology)) | Idea or Concept |
| Epidemiological Factors | Idea or Concept |
| epidemiological indicators | Qualitative Concept |
| Epidemiological Methods | Research Activity, Intellectual Product |
| Epidemiological Monitoring | Occupational Activity |
| Epidemiological Studies (Epidemiologic Studies) | Research Activity |
| Epidemiological Studies (Epidemiology Research) | Research Activity |
| Epidemiological Techniques (Epidemiological Methods) | Research Activity, Intellectual Product |

If you did not find the concept you're looking for, and you're sure your spelling is correct, please click:

Suggest New Term

Center
for
Bioinformatics

FIRSTGOV
Your First Click to the U.S. Government

Search

The basic search is enabled. Click on the "Advanced" tab to customize your search criteria. You can mouse-over each advanced search item for help in utilizing the advanced features.

CL003599: Epidemiological Factors

Idea or Concept

Epidemiological Factors Definitions

| Source | Definition |
|--------|---|
| NCI | those things which contribute to the establishment and transmission of disease. |

Epidemiological Factors Synonyms

- Epidemiological Factors

Epidemiological Factors Sources

NCI

View neighborhood in

Table View Of Sources

Concepts More General Than Epidemiological Factors

(inverse isa) Conceptual Entities

Concepts More Specific Than Epidemiological Factors

- (isa) Chemical Agents
- (isa) Dietary Factors
- (isa) Endogenous Factors
- (isa) Epidemiology, Age, Sex
- (isa) Epidemiology, Geographic Location
- (isa) Epidemiology, Occupation
- (isa) Epidemiology, Race/Ethnic Origin
- (isa) Epidemiology, Tobacco, Alcohol, Drugs
- (isa) Exogenous Factors
- (isa) Knowledge, Attitudes, Behaviors
- (isa) Lifestyle Factors, Unspecified
- (isa) Societal Factors

Concepts Related To Epidemiological Factors

Medications

Procedures

Laboratory

Diagnosis

NCI Metathesaurus



DIVISION OF
CANCER CONTROL
& POPULATION SCIENCES

[Basic](#)
[Advanced](#)

The basic search is enabled. Click on the "Advanced" tab to customize your search criteria. You can mouse-over each advanced search item for help in utilizing the advanced features.

[Concept](#) | [Definitions](#) | [Synonyms](#) | [Sources](#) | [Broader Concepts](#) | [Narrower Concepts](#) | [Related Concepts](#) | [Medications](#) | [Procedures](#) | [Laboratory](#) | [Diagnosis](#) | [Open NCI Hierarchy](#) | [View Hierarchy Location](#)

CL000367: Epidemiology, Tobacco, Alcohol, Drugs

Research Activity

Epidemiology, Tobacco, Alcohol, Drugs Definitions

| Source | Definition |
|--------|--|
| NCI | consideration of tobacco, alcohol, or drug use as variables in disease transmission and control. |

Epidemiology, Tobacco, Alcohol, Drugs Synonyms

- Epidemiology, Tobacco, Alcohol, Drugs

Epidemiology, Tobacco, Alcohol, Drugs Sources

View neighborhood in

[Table View Of Sources](#)

Concepts More General Than Epidemiology, Tobacco, Alcohol, Drugs

(inverse_isa) [Epidemiological Factors](#)

Concepts More Specific Than Epidemiology, Tobacco, Alcohol, Drugs

Concepts Related To Epidemiology, Tobacco, Alcohol, Drugs

Medications

Procedures

Laboratory

Diagnosis



NCI Metathesaurus



DIVISION OF
CANCER CONTROL
& POPULATION SCIENCES



Basic

Advanced

Search

The basic search is enabled. Click on the "Advanced" tab to customize your search criteria. You can mouse-over each advanced search item for help in utilizing the advanced features.

Sources of Epidemiology, Tobacco, Alcohol, Drugs

| Source | Type | Code | Term |
|---------------|------|-------------------------|---------------------------------------|
| NCI | PT | C18774 | Epidemiology, Tobacco, Alcohol, Drugs |
| RAEB-2 | PT | N023242 | Epidemiology, Tobacco, Alcohol, Drugs |

An explanation of the UMLS term types can be found at the NLM's documentation page [here](#).

About
Browse
Copyright
New Term
Sources
User's Guide

